A Method Of Detecting Gravitational Wave Based On Time-frequency Analysis And Convolutional Neural Networks

Xiangru Li,^{1,*} Woliang Yu,¹ and Xilong Fan^{2,†}

¹School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China ²School of Physics and Electronics, Hubei University of Education, Wuhan 430205, China

This work investigated the detection of gravitational wave (GW) from simulated damped sinusoid signals contaminated with Gaussian noise. We proposed to treat it as a classification problem with one class bearing our special attentions. Two successive steps of the proposed scheme are as following: first, decompose the data using a wavelet packet and represent the GW signal and noise using the derived decomposition coefficients; Second, detect the existence of GW using a convolutional neural network (CNN). To reflect our special attention on searching GW signals, the performance is evaluated using not only the traditional classification accuracy (correct ratio), but also receiver operating characteristic (ROC) curve, and experiments show excellent performances on both evaluation measures. The generalization of a proposed searching scheme on GW model parameter and possible extensions to other data analysis tasks are crucial for a machine learning based approach. On this aspect, experiments shows that there is no significant difference between GW model parameters on identification performances by our proposed scheme. Therefore, the proposed scheme has excellent generalization and could be used to search for non-trained and un-known GW signals or glitches in the future GW astronomy era.

PACS numbers: 98.80.Es, 95.30.Sf, 95.85.Sz

1 INTRODUCTION

Gravitational waves (GWs) potentially gives us a remarkable opportunity to see the very early universe and enables us penetrate unprecedented regions of universe and compact astronomical objects. Particularly, the direct detections of GWs by advanced LIGO detectors and advanced Virgo [1–5] further enhance people's interests in gravitational theory and its observations and have brought us into the gravitational-wave astronomy era [6–9].

Burst pipelines used in LVC (LIGO-Virgo Collaboration) search for generic gravitational-wave transients with minimal assumptions about the signal waveform, polarization, source direction, or time of occurrence [10]. Although main targets of burst pipelines are un-modeled transient gravitational wave events, the CBC (compact binary coalescence) signal can also be identified by burst pipelines [11]. The characteristic parameters of these un-modeled GW signals, such as the central frequency, duration and bandwidth, can be estimated from the time-frequency map of data and then converted to real physical parameters by comparing with more sophisticated waveforms [11].

This work focuses on the procedure identifying the existence of simulated damped sinusoid GW signals from data with Gaussian noise. This type waveforms could be a burst of gravitational waves from the ringdown of a perturbed black hole[12] which could be the remnant BNS (binary neutron star) post-merger GWs, and excitation of fundamental modes in neutron stars [13]

Recently, convolutional neural network (CNN) based data analysis methods are developing to analyze the

time-domain GW data [34, 35]. Studies of this work show that discriminant characteristics of GW signal time series are more evident in time-frequency map. Therefore, we propose to decompose the data using a wavelet packet (WP) [14–16] and represent it by the derived decomposition coefficients. Based on this WP representation, the existence of GW is identified using a CNN. The CNN can tolerate the uncertainty of GW emerging time in an observed data stream.

2 DATA

Let $s \in \mathbb{R}^D$ represent an observed signal, $y \in \{1, 0\}$ is its label: y = 1 indicates that there exists a GW signal in s, and y = 0 indicates that there isn't any GW signal in s. This work is to establish a model

$$y = f(s), \tag{1}$$

using statistical theories to detect the existence of GW component in our observed signal. This is referred to as a classification problem in machine learning community.

In this kind of statistical classification schemes, the knowledge of GW signal detection is expressed using a set, $S_{tr} = \{(s_i, y_i), i = 1, \dots, N\}$, of observed/theoretical signals and their labels, where $s_i \in R^D$ an observed signal, $y_i \in \{1,0\}$ is its label indicating whether there exists GW signal in $s_i \in R^D$. That is to say, the empirical data S_{tr} are knowledge carriers. We should estimate the model y = f(s) from this empirical data set S_{tr} using some computational schemes. For conveniences, this estimation is denoted by \hat{f} . Therefore, this empirical set S_{tr} is referred to as a training set in

related communities. The empirical signal (s, y) is called a sample.

To evaluate the performance of \hat{f} , some samples independent of S_{tr} are also needed. These evaluation samples constitute a set S_{te} , referred to as test set. The independence here is to ensure the objectivity of the model evaluation.

One typical characteristic of the problem is that the GW is usually very weak and the observed signal is contaminated with various disturbances. All of the existing disturbances are collectively referred to as noises. This work assumes that the glitches are removed from the data or from the detector entirely [17–24]. Therefore, we modeled the noises approximately using a Gaussian distribution with zero mean $N(0, \sigma^2)$, where σ^2 is the variance of the Gaussian distribution. Let $s \in R^D$ is an observed signal, and h and h are the GW component and noises in h Since the proposed scheme based on wavelet packet and CNN and its characteristics are totally different from the matched filter techniques in related literatures, the ratio of maximum amplitude of the waveform and the white noise standard deviation

$$A^{S/N} = \frac{\max(\mathbf{h})}{\sigma(\mathbf{n})},\tag{2}$$

is defined to evaluate the performance of our approach as a scale of the data quality, rather than optimal signal-to-noise ratio (SNR) [35–37]. For an easy comparison, some experimental results are evaluated using both $A^{S/N}$ and the optimal SNR in this work.

In this work, the GW signal is computed using damped sinusoid waveform [25]:

$$\mathbf{h}(t) = e^{-\frac{t-t_0}{\tau}} \sin(2\pi f_0(t-t_0)) \tag{3}$$

where t_0 is the emerging time of the GW signal, f_0 is the central frequency and $\tau = Q/(\sqrt{2}\pi f_0)$ is the decay parameter. The frequency f_0 covers a range [40, 200] Hz uniformly, and quality factors Q randomly takes three values 3, 9 and 100 [25] as the *fiducial* models in simulated data. In this paper, the detector antenna function and distance effects are hidden in the amplitude ratio assumption.

In a supervised learning scheme, data are split into two separated sets respectively for training and testing. Our fiducial training data set contains 25000 only Gaussian noise and 25000 signals injected in Gaussian noise. Our fiducial test data set contains 95000 only Gaussian noise and 95000 signals. In application, it is difficult to ensure that the $A^{S/N}$ of every test data to be processed is in the range of training data. Therefore, the test data are computed with a broader $A^{S/N}$ range than that $A^{S/N}$ in training data. In this work, the $A^{S/N}$ ranges of training data and test data are [0.2, 0.8] and [0.15,1.05] respectively. We also simulate 95000 test signals with Q = 30, 50, 70 to test the generalization of our approach.

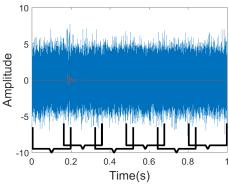


FIG. 1: A sample of GW. In this Figure, the blue curve is a signal to be analyzed, which consists of GW signal and noise. The red curve is the theoretical GW signal. The $A^{S/N}$ coverage of this work is [0.15,1.05]. For visibility, this figure shows a sample with $A^{S/N}=0.3$. In application, signal is obtained in stream. The GW detection system needs to move the time window on the stream to take the signal segment in it to analyze. This diagram shows our proposed schemes for moving time windows (indicated with bracket). Two other parameters of this signal are $f_0{=}100~{\rm Hz},~{\rm Q}{=}9$.

Each data set is simulated with a different Gaussian noise realization for both training and test data set. A signal with $A^{S/N}$ =0.3 is presented in Fig. 1.

In applications, the signal is obtained continuously in a form of data streams. The GW signal detection system needs to move a time window on the data stream, extracts a segment of the signal from every window, and determines whether there exist a GW signal (Fig. 1) in the signal segment. This work propose a way to move the detection window on the data stream with some overlapping to increase identification performance. A segment of the signal from every window is an above-mentioned sample \boldsymbol{s} .

3 DETECTION SCHEME



FIG. 2: A Flowchart to show the order of procedures in the proposed scheme.

3.1 Frame

Theoretically, we could directly use the model in equation (1) to determine whether there exists a GW signal component in s. However, it is shown that there is a

clear difference between a GW signal and noises (Fig. 1) in frequency space (Fig. 3). Therefore, we can project the signal s to be analyzed into a frequency space,

$$z = g(s) \tag{4}$$

before determining whether there exist a GW component in waves in \boldsymbol{s} as follows

$$y = h(z) = h(g(s)). \tag{5}$$

The z is the representation of s in a frequency space. In this work, function g is a WP decomposition, and function h is estimated using a CNN network from a training set.

Actually, scheme (5) is a special implementation of model (1) by splitting it into two sequential procedures: Firstly, preprocess the observed signal s by z = g(s), then determine the existence of GW signal component in s by y = h(z). This splitting scheme can simplify a complex work to some degree. In machine learning and data mining communities, the procedure z = g(s) is called feature extraction. The feature extraction is to simply the establishing of model (5) by finding a more approximate representation of the data to be analyzed, s here, and removing irrelevant or weakly-related data components. A flow chart of the proposed scheme is presented in Fig. 2. We will elaborated it in the following parts of this section.

Typical frequency analysis schemes are Fourier transform [26], Wavelet transform and WP transform [14– 16]. Wavelet transform and WP transform have a timefrequency analysis capability, while the Fourier transform has no time analysis capability. Our experiments show that WP is applicable in the GW detection problem and there isn't significant difference on GW identification between typical Wavelet bases db1, bior1.3, coif4, rbio4.4, sym4 and haar. Therefore, we propose to use WP transform with db1 base. A brief introduction to the principle and implementation of WP analysis can be found in [27]. This introduction need very little mathematical knowledge to read. For WP has both time analysis and frequency analysis capabilities, the decomposition of an observed signal (Fig. 1) is an image in a two dimensional space (Fig. 3).

3.2 Deep Convolutional Neural Networks

One typical characteristic of the GW detection problem is that the emerging time of GW signal is uncertain in the detection window (Fig. 1). This characteristic is called translation invariant in machine learning and computer vision community. After WP decomposition, a signal to be analyzed is represented with a two-dimensional image with translation uncertainty (Fig. 3). For this kind problem, a typical scheme is CNN network [28, 29].

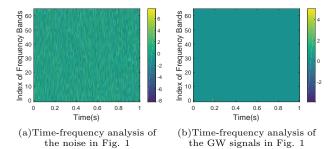


FIG. 3: Comparing the noise and GW signals in time-frequency space. In this figure, color indicates the intensity of a signal. The horizontal axis is time, and vertical axis is the index of frequency bands in WP decomposition. The frequency in this figure and that of the GW signal are different although there are positive correlation between them.

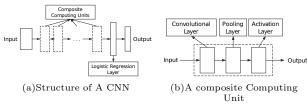
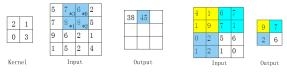


FIG. 4: Structure of a convolutional neural network (CNN). (a) The overall structure of a CNN network; (b) Structure of a composite computing unit (CCN). In this work, we used two CCUs. Therefore, two convolutional layers were used (Fig. 4(b)). Every convolutional layer consists of a series of convolutional kernel (Fig. 5(a)).

The CNN has been widely used in image understanding and computer vision [30].

The structure of the CNN used in this wok is presented in Fig. 4. This CNN network consists of an input layer, several composite computing units (CCU) modules, and a LOGISTIC regression layer (also called an output layer) (Fig. 4(a)). The input layer receives the WP decomposition (Fig. 3) of a signal to be processed (Fig. 1). A CCU module consists of a convolutional layer, a pooling layer, and an activation layer (Fig. 4(b)). This work used a CNN with two CCU modules. Experiments shows this CNN can detect signal features with different scales.

In a CNN network, there are two key concepts, convolution kernel (CK) and pooling. A convolution kernel is a matrix (the left subfigure in Fig. 5(a)) which characterizes a discriminant pattern of the signals in the problem to be investigated. In a convolution layer, we move the CK on the signal to be analyzed, make convolution operation on CK and the signal segment under CK. By the convolution operation, we can evaluate the presence of some discriminant characteristics. The results of computational operation is referred to as convolution response. The input into a convolution layer is the output from its previous layer of the neural network. The previous layer can be the input layer of the CNN and an activation



(a)Convolutional operation

(b)An example of max pooling

FIG. 5: Two sketch maps of convolutional operation and max pooling. (a) The output 38 is computed as $3\times5+0\times7+1\times7+2\times8=38$ by aligning the reversed kernel with input at 3 and 5, output 45 is computed as $3\times7+0\times6+1\times8+2\times8=45$ by aligning the reversed kernel with input at 3 and 7. (b) Max pooling is to obtain an output by applying a max filter to (usually) non-overlapping subregions of the initial representation: output 9 is computed by maximizing 4, 1, 1, 9 (yellow part), 7 computed by maximizing 6, 7, 7 and 1, and so on. The input in Fig. 5(a) is a time-frequency image (Fig. 3). This work used two convolution layers (Fig. 4(b)) with 15 convolution kernels and 20 convolution kernels respectively.

layer (Fig. 4). Therefore, selections of the convolution kernel has fundamental influence on the performance of the CNN network, and depends on the characteristics of the problem to be investigated. For the knowledge of GW signal is embodied in training data in this kind statistical schemes, configurations of the convolution kernel are learned from training data. The learning method is back propagation algorithm [31].

In the results of convolution operation, there is usually much redundancy. The existence of redundancy can result in evident degradation of detection performance on test data. This phenomenon is called as over-fitting in machine learning. To overcome this kind over-fitting problem, an operation, pooling, is adopted. The pooling operation reduces the redundancies from data by merging the convolution responses in every pooling window (Fig. 5(b)). The merging is implemented by computing the maximum of the convolution responses in a merging window in this work. This merging window is a 4×4 matrix in this work. More about pooling can be found in [32] and [33].

4 EXPERIMENTAL EVALUATIONS

4.1 Learning

The structure of the GW signal detection system is introduced in Fig. 2 and Fig. 4. In this system, the following parameters need to be determined: number of composite computing units, number of convolution kernel in every convolution layer, size and configurations of every convolution kernel, size of pooling window and the parameters in LOGISTIC REGRESSION layer.

Theoretically, the above-mentioned parameters can be selected based on some optimization theories using training data. However, it is a hybrid, complex optimization

problem consisting of discrete parameters and continuous parameters. The number of every composite computing units, the number of convolution kernels in every convolution layer, size of every convolution kernel and size of pooling window are discrete should be positive integers and are discrete parameters. And configuration of every convolution kernel and the parameter of LOGISTIC REGRESSION are from real space and continuous parameters. Therefore, it is a complex and hybrid optimization problem.

To make its computational complexity be acceptable, the discrete parameters are chosen based on experiences. We used two composite computing units, numbers of convolution kernels are 15 and 20 respectively for the two convolutional layers from input end to output end of the CNN (Fig. 4(a)). The continuous parameters are estimated using back propagation algorithm [31] from training data .

4.2 Performance Evaluation

To evaluate the proposed recognition scheme, we utilized five typical measures precision, recall, accuracy and a receiver operating characteristic (ROC) curve. Suppose S is a sample set of observational signals, $S_1 = \{(s,y): (s,y) \in S \text{ and there exists GW in } s\}$, $S_2 = \{(s,y): (s,y) \in S \text{ and } (s,y) \not\in S_1\}$, \hat{S}_1 and \hat{S}_2 are respectively the estimation of S_1 and S_2 based on a recognition system. Then, the precision and recall on S are defined as following:

$$\operatorname{Precision}(S) = \frac{|S_1 \cap \hat{S}_1|}{|\hat{S}_1|},\tag{6}$$

$$\operatorname{Recall}(S) = \frac{|S_1 \cap \hat{S}_1|}{|S_1|},\tag{7}$$

$${\tt Accuracy}(S) = \frac{|S_1 \cap \hat{S}_1| + |S_2 \cap \hat{S}_2|}{|S|}, \tag{8}$$

That is, the precision measures the reliability of alarm from a GW recognition scheme (the detection probability in GW community language), recall evaluates detection completeness of the proposed scheme, and accuracy measures the performance on the whole.

In this work, there are two class of samples: one consists of noise and GW signal, the other is noise. For convenience, these two classes are denoted with NG and N, respectively. Therefore, detecting of GW is equivalent with classifying a sample into NG or N. For this kind of problem, a typical performance evaluation method is the above-mentioned accuracy.

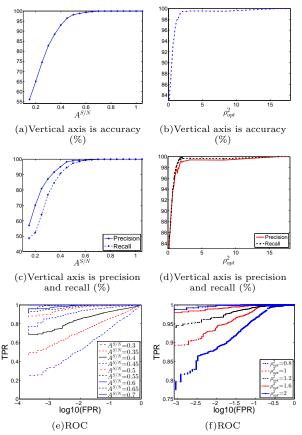


FIG. 6: Dependencies of detection performance on SNR. To facilitate reading, the work evaluates the performances using both $A^{S/N}$ and the optimal SNR ρ_{opt}^2 [35–37].

However, recognition difficulties may depend on $A^{S/N}$. Fig. 6(b) shows the dependencies of detection accuracy on $A^{S/N}$. On the whole, this scheme has excellent detection accuracy: the detection accuracy is above 96% on data with $A^{S/N}$ low to 0.45, above 82% even for data with $A^{S/N}$ low to 0.3.

In GW detection, our interests on these two class is unbalanced. We care more about the recognition performance for GW than that for the other one. For example, what percentage of the GW signal can be detected successfully, and how many of the samples classified as GWS are correct? The precision and recall are two measures scoring these unbalanced interests. Experimental results in Fig. 6(d) show that precision and recall are above 98% and 94% on data with $A^{S/N}$ low to 0.45, above 87% and 76% even for data with $A^{S/N}$ low to 0.3.

Comparing with the related research results in [34], an eminent characteristic of the proposed scheme is its robustness to noise. The results in Fig. 6 show that the accuracy is higher than 88% if $A^{S/N}>0.4$, accuracy higher than 96% if $A^{S/N}>0.5$, and rapidly approaches to 99% if $A^{S/N}>0.65$. Therefore, the performance of the proposed scheme is excellent for low $A^{S/N}$ data, which is the case in gravitation wave detection. These two works

of [34] and this paper are carried out all using CNN. The differences are that this paper transformed the data using WP transform and identify the existence of GW signal in time-frequency space, however, George and Huerta [34] do the identification in time space.

The ROC evaluation results are presented in Fig. 6(f). In this figure, we use another name True Positive Rate (TPR) for recall to be consistent with literatures using ROC curve. False Positive Rate (FPR)

$$FPR = \frac{|\hat{S}_1 \cap S_2|}{|S_2|}.\tag{9}$$

Actually, FPR is an indicator of false alert (false-alarm probability in GW community language). A small FPR means a ROC curve close to True positive rate axis, and a high precision/TPR means a ROC curve close to line TPR=1. The results in Fig. 6(f) show that the ROC curve quickly approach to the True positive rate axis and line TPR=1 in case of $A^{S/N}$ increase from 0.3 to 0.7. Therefore, the performance of the proposed scheme is very acceptable.

4.3 Experimental Evaluations: Efficiency and Generalization

The proposed scheme consists of two procedures: decompose a signal using wavlete packet, and identify the existences of a GW signal. These two procedures averagely take 16.21 milliseconds to process a sample (equivalently 16.21 seconds/thousand sample) on a HP computer using two i5-3470 CPUs with frequency 3.2GHz.

In theoretical model (3), parameter $\tau = Q/(\sqrt{2}\pi f_0)$ controls the decay of a GW signal. A larger τ makes the GW signal decrease more rapidly and be weaker relatively to noise. Learning of the proposed scheme uses the information with Q=3,9,100. Can this scheme identify the GW signal with different Q? By generalization, we refer to this problem. We test the generalization of our proposed scheme by searching for non-trained signals with Q=30,50,70 with the CNN trained with fiducial models (Q=3,9,100). As shown in Fig. 7, there is no significant difference between different Q values on identification performances. Therefore, the proposed scheme has excellent generalization.

5 SOME TECHNICAL PROBLEMS AND TRANSFERABILITY OF THE PROPOSED SCHEME

Section discusses the optimal configuration problem of the proposed scheme and its transferability.

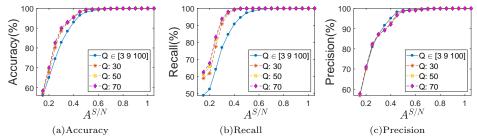


FIG. 7: Performances of the proposed scheme on general non-trained signals with different Q.

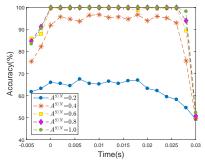


FIG. 8: Selection of detection window scheme. Horizontal axis is the emerging time of GW signal (Equation (3)).

5.1 Optimization of detection window

As discussed in Section and Fig. 1, signals are obtained in data streams in application, and our GW signal detection system needs to move a time window on the stream to get data sample for determining the existences of GW. And there are two ways to move ways to move the time window. One way is to using non-overlapping window scheme, and the other using overlapping windows (Fig. 1).

To make choice between these **two time window** schemes, we conducted an experiment by testing the dependency of detection performance on GW wave emerging time (Equation 3) in the time window. The experimental results are presented in Fig. 8. In this experiment, the time window width is 0.03s (the segment [0, 0.03] on the horizontal axis). It is shown that, if a GW signal emerges before 0.026s, the detection performance is steady on a high level. While if it emerges at the end of time window (later than 0.004s), the performance will decrease evidently. Therefore, we take the overlapping time window scheme in Fig. 1 with overlapping width 0.004s. By doing this, every GW signal can emerge winthin the high performance area in one detection window.

Actually, the determination of overlapping window width depends on characteristics of gravitational wave (GW) signal. As being stated in section , this work considers the GW signals on frequency range [40, 200]Hz. On this frequency range, the strongest peak appears within 0.004 seconds after emerging time t_{start} (equation 3) for

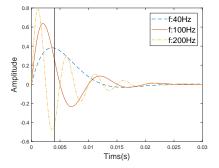


FIG. 9: Three theoretical gravitational wave signals. The strongest peak appears in 0.04 seconds for every presented signal. f: frequency.

every GW signal (Fig. 9). Therefore, the experimental results in Fig. 8 essentially indicate that the information from the strongest peak play an important role in GW signal detection. It also tell us that our overlapping window width should be adjusted correspondingly in case of different GW frequency range.

Another problem is how to determine an acceptable detection window width (DWW). On this aspect, we conducted several experiments with variable window widths. The experimental results are presented in Fig. 10. The experimental results show strong sensitivity of detection performance to DWW. On the whole, the detection performance decreases when DWW increase. However, the characteristics of performance sensitivity to DWW depend on the SNR of the GW signal to be detected. There should be a balance between GW signal with high SNR and GW signals with low SNR. In this work, we choose DWW =0.03 second. In applications, our optimal DWW can be selected based on the SNR range and percentage of low SNR GW signal.

5.2 On signal resolution

A signal is represented using a series of energy values sampled evenly based on the time axis in this work. Signal resolution (SR) refers to time interval between two successive sampling points. The experiments in Fig. 11 investigate the sensitivity of detection performance to SR on data sets with variable DWWs. It is shown that a high

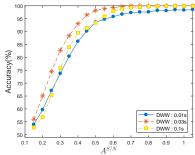


FIG. 10: Influences of detection window width (DWW) on performance.

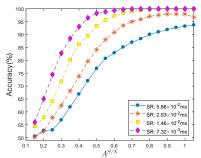


FIG. 11: Effects of signal resolution (SR) on detection performance.

SR, low time interval between two successive sampling points, means good detection accuracy on the whole.

However, a high SR means that we have more energy values for each patch of signal, a sample, on a specific detection window and we should compute and analyze in a high dimensional space. Therefore, a high SR, with large time interval between successive sampling points, is closely related with a high computational complexity and we should keep a balance between detection accuracy and computational complexity in applications. This work use the $SR = 7.32 \times 10^{-3}$ milliseconds (equivalently 136533 pixels per second).

5.3 On frequency analysis methods

A signal is represented using a series of energy values sampled evenly based on a time axis in this work (Section 2, Fig. 1). Based on analyses of the frequency characteristics of GW signal (Fig. 3), we propose to represent a signal based on a frequency analysis method, Wavelet Packet. Actually, there are other options for frequency analysis, for example, Wavelet transform, Fourier Transform. Why do we choose Wavelet Packet from them?

Therefore, we conducted some experiments in Fig. 12 to evaluate the performances of these three types representations: energy values (original signal), Wavelet Transform, Wavelet Packet and Fourier Transform. It is shown that the original signal achieves the poorest performance among these four representations, and Wavelet Transform and Wavelet Packet acts evidently better than

the other two on the whole (Fig. 12(a)). Althogh Wavelet Transform achieve an accuracy and a precision slightly better than those of Wavelet Packet(Fig. 12(a) and Fig. 12(b)), Wavelet Packet has a evidently better recall on the whole (Fig. 12(c)). That is to say, the detection scheme based on Wavelet Packet miss GW signal much less, which is very important for this problem. Therefore, we propose to use Wavelet Packet in detecting GW signals.

5.4 Evaluations of pooling schemes

The proposed scheme consists of two stages: frequency analysis and GW detection. The GW detection is implemented using CNN network. In CNN network, a key procedure is pooling. And two typical pooling methods are max pooling and average pooling (also referred to as mean pooling in some literatures). The experimental results in Fig. 13 show that max pooling method achieves a significantly better performance in GW signal detection, even if the SNR as low as 0.15.

Actually, these experimental results are consistent with the characteristics of GW signal. Fig. 3 shows that the information is localized in a small area in time-frequency space. For this kind of signals, the max pooling scheme can be more easier to detect the appearance, while the average pooling is prone to mixing the GW peak information with the data components nearby.

6 CONCLUSIONS AND DISCUSSIONS

In this work, we propose a general scheme for GW signal detection. This scheme has excellent detection accuracy: the detection precision is above 98% on data with $A^{S/N}$ low to 0.45, above 87% even for data with $A^{S/N}$ low to 0.3.

The proposed general scheme consists of two procedures: firstly, extract features using WP; secondly, detect the existence of GW signal using a CNN. It is shown that WP spotlights characteristics of GW signal in time-frequency space, and can improve detection performance. In reality, the observed signal comes as a data stream. This work propose an overlapping window scheme, by which we reformulate the GW signal detection problem as a classification issue in machine learning.

The experiments on Q, a parameter in GW theoretical model, shows that the proposed scheme has excellent generalization capability. Generalization performance is a meansure of detection capability on observations unknown to learning stage. In application, it is inevitable that there is some variations or disturbances in the obtained GW signals. Therefore, this is crucial for searching un-modelled signals with any machine learning approach which is based on training data. Actually, this is

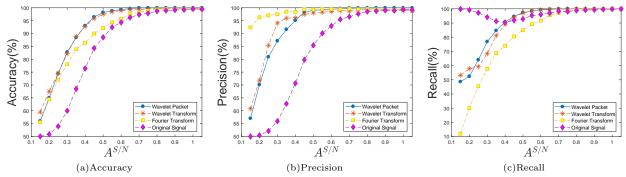


FIG. 12: Selection of frequency analysis methods. In this investigation, the lowest SNR is 0.15, Signal resolution is $SR = 7.32 \times 10^{-3}$ milliseconds (equivalently 136533 pixels per second) and the overlapping width of the detection window is 0.004 seconds.

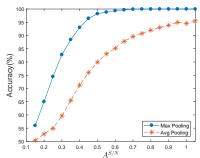


FIG. 13: Selection of pooling scheme. In this investigation, the lowest SNR is 0.15.

a very flexible scheme as a reference for related scientists in this field. For example, if preparing some training samples for one or more classes of glitches and replacing the LOGISTIC regression with SOFTMAX regression [32], the proposed scheme can be explored in glitch identification/detection; if replacing the LOGISTIC regression with an linear regression, we can study the application of the proposed scheme in GW parameter estimation in theory, for example, f_0 and τ in equation 3.

However, this proposed scheme cannot directly extended to estimating parameter t_0 in GW model of equation 3, the emerging time of a GW signal in data steam. The result is caused by the pooling operation in CNN. Therefore, we at least replace the CNN procedure with another regression method to investigate novel t_0 estimation scheme.

Acknowledgments.— Authors are grateful for valuable suggestions and corrections from Dr. Jin Li. X. L. and W. Y. are supported by the National Natural Science Foundation of China (grant No: 61273248, 61075033), the Natural Science Foundation of Guangdong Province (2014A030313425, S2011010003348). X. F. is supported by Natural Science Foundation of China under Grants (No. 11633001, No. 11673008) and Newton International Fellowship Alumni Follow on Funding.

- * Electronic address: xiangru.li@gmail.com
- † Electronic address: fanxilong@outlook.com
- B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 116,061102 (2016).
- [2] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X 6, 041015 (2016)
- [3] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 118, 221101 (2017)
- [4] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 119, 161101 (2017)
- [5] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 119, 141101 (2017)
- [6] S. Adrián-Martínez et al. (ANTARES, IceCube, LSC and Virgo), Phys. Rev. D 93, 122010 (2016)
- [7] B. P. Abbott et al. Astrophys. J. Lett. 848, L12 (2017)
- [8] B. P. Abbott et al. (LSC, Virgo, Fermi-GBM, and IN-TEGRAL), Astrophys. J. Lett. 848, L13 (2017)
- [9] B. P. Abbott et al. (LSC and Virgo) plus six astronomy teams, Nature, 2017
- [10] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D. 95, 042003 (2017)
- [11] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 93, 122004 (2016)
- [12] C. Vishveshwara, Nature 227, 936 (1970)
- [13] O. Benhar, V. Ferrari, and L. Gualtieri, Phys. Rev. D 70, 124015 (2004).
- [14] S. Mallat, A Wavelet Tour of Signal Processing (3rd ed.), Boston: Academic Press) (2009)
- [15] S. Mallat, IEEE Trans. on Pattern Analysis and Machine Intel., 11(7), 674 (1989).
- [16] I. Daubechies, Ten Lectures on Wavelets, Philadelphia: Society for Industrial and Applied Mathematics (1992).
- [17] J. Powell, D. Trifir, E. Cuoco, et al. 2015, Classical and Quantum Gravity, 32(21): 215012.
- [18] J. Powell, Jade, A. Torres-Forné, R. Lynch, et al. Classical and Quantum Gravity 34.3: 034002 (2017).
- [19] D. George, H.Y. Shen, and E.A. Huerta arXiv:1706.07446 (2017).
- [20] Zevin, Michael, et al. 2017, Classical and Quantum Gravity, 34.6: 064003.

- [21] N. Mukund, S. Abraham, S. Kandhasamy and N.S. Philip, Phys. Rev. D, 95.10:104059 (2017).
- [22] Bose, Sukanta, et al. 2016, Phys. Rev. D, 94.12: 122004.
- [23] S. Bahaadini, N. Rohani, S. Coughlin, et al. 2017, arXiv:1705.00034.
- [24] Massimiliano Razzano and Elena Cuoco. Image-based deep learning for classification of noise transients in gravitational wave detectors, https://pnp.ligo.org/P1700254/
- [25] J. Abadie, B.P. Abbott, R. Abbott, et al. Phys. Rev. D, 85, 122007 (2012).
- [26] K.B. Howell. Principles of Fourier analysis[M]. CRC Press (2016).
- [27] X.R. Li, Yu Lu, G. Comte, et al. Astrophys. J. Suppl. S., 218(1),3 (2015).
- [28] Y. LeCun, B.E. Boser, J.S. Denker, et al. In Advances in neural information processing systems, 396 (1990)
- [29] Y. LeCun, L. Bottou, Y. Bengio, Y., & Haffner, P., Proceedings of the IEEE, 86, 2278 (1998).

- [30] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Nature, 521, 436, 2015.
- [31] D.E. Rumelhart, G.E. Hinton, R.J. Williams, 1986, Nature. 323 (6088): 533536 (1986)
- [32] Andrew Ng, J. Ngiam, C.Y. Foo, et al. Unsupervised Feature Learning and Deep Learning. http://ufldl.stanford.edu/tutorial/ (2013).
- [33] Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning. MIT Press (2016).
- [34] D. George, E.A. Huerta, arXiv:1701.00008v1 (2017).
- [35] Daniel George, E. A. Huerta, arXiv:1711.03121 (2017)
- [36] B. J. Owen and B. S. Sathyaprakash, Physical Review D 60, 022002 (1999).
- [37] Hunter Gabbard, Fergus Hayes, Chris Messenger, and Michael Williams. Matching Matched-Filtering with Deep Networks for Gravitational wave Astronomy, in preparation.